



Free Training Transcript:
Data Prep with Text
and Excel Files

Welcome to this Data Prep video. You can download the dataset to follow along in your own copy of Tableau

Here we have a report in Excel, showing the number of resolved incidents per Employee per month. Formatted reports like this aren't easy to do analysis on. The "Ideal" tab shows data formatted like a database table, or raw data. This is the format Tableau prefers for optimal analysis. However, there are several features in Tableau Desktop to help automatically reshape Text and Excel files into this format for better analysis in Tableau. To begin, we'll connect to this Excel file. In Tableau Desktop select Excel from the Connect Pane. Navigate to the downloaded file and click open. We'll drag out the "Resolved Incidents" sheet.

Data Interpreter

Although Tableau can connect to this report, we can see here in the preview that this isn't a great format. There are no column names, the headers from Excel have a lot of nulls, and so on. In the side panel, we see the option to Use the Data Interpreter. Let's click turn on. Now we see that those headers and nulls have been stripped out, and our columns are properly identified. We could go on, but if we want to get more specifics on what the Data Interpreter did, we can click "Review the Results". This will open an Excel file describing the changes. If we click to the tab we used, Resolved Incidents, we see which fields are being used as headers, in red, and which are considered data, in green.

Before we go back to Tableau and our data source, let's take one more look at that "Ideal" tab. Note that instead of having a column for each month with data underneath, in this format, there is a "Date" column and each row contains the number of resolved incidents for each unique combination of date and employee. This data is taller, with more rows, rather than wider, with more columns.

Pivot

Back in Tableau, we want to change the format from that column-per-month layout into a single date column and a single column for Resolved Incidents. To do this easily, we'll simply select all the date columns. Click on the first, scroll if necessary, then shift click on the last. We'll open the menu and select "Pivot". This pivot feature essentially merges the information from the original columns and rows into two new columns – Pivot field names, and Pivot field values. We can see that "Pivot field names" is actually our Date, and "Pivot field values" can be renamed "Resolved Incidents". We can double click to rename.

Metadata Grid

If we click on this icon here above the first column, we're brought to the metadata grid. This can be a useful view, as the vertical layout can be easier to navigate, especially if you have a large number of fields, and it's also useful when tables have been joined, or to see the original field names. For now, we'll go back to the standard view.

Split

Our data is almost where we want it to be, but there's one more thing we can do. Here, this "Employee" field is actually two pieces of information – a location code, A, B, C, D, or E, followed by an Employee ID number. We can split the column based on the shared delimiter of that hyphen. Click to open the menu and select Split. We'll use the Metadata Grid view and rename our split fields. Split 1 should be Location, and, Split 2 should be Employee ID. There's an Abc next to the Date field indicating this column is considered a String. We know it's actually a Date, though, so we can click on the Abc and select Date to update the data type. Now if we click on Sheet 1, we'll see nice tidy data ready for analysis!

Custom Split

Let's bring Employee ID to the view, Resolved Incidents to Columns. When we sort it's clear that there are really two groups of employees – some who resolve a much higher number of incidents than others – those who are tier II. If we add another data source, connecting to that same excel file, we see there's a tab called Tiers. This report adds a –II to the end of an employee ID if they're tier II. Because not all rows have this ending, a standard split won't work. However, we can use a Custom Split – open the menu and select Custom Split. We can choose our delimiter, we'll use a hyphen. And mandate that we want to break off all columns. This forces Tableau to break off that 3rd column with the tier II indicator.

Conclusion

Thank you for watching this Data Prep training video. We invite you to continue with the Free Training videos to learn more about using Tableau.